



DOT/FAA/AR-00/43

Office of Aviation Research
Washington, DC 20591

Test and Evaluation Plan for the Laboratory Validation of X-Ray Threat Image Projection

Susan B. Monichetti
Michael D. Snyder

Aviation Security Research and Development Division
Federal Aviation Administration
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

May 2000

This report is approved for public release and is on file at the William J. Hughes Technical Center, Aviation Security Research and Development Library, Atlantic City International Airport, New Jersey 08405

This document is available to the public through the National Technical Information Service (NTIS), Springfield, Virginia, 22161



REPRODUCED BY:
U.S. Department of Commerce
National Technical Information Service
Springfield, Virginia 22161

NTIS

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof. The United States Government does not endorse products or manufacturers. Trade or manufacturer's names appear herein solely because they are considered essential to the objective of this report. This document does not constitute FAA certification policy.

1. Report No. DOT/FAA/AR-00/43		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle Test and Evaluation Plan for the Laboratory Validation of X-Ray Threat Image Projection				5. Report Date May 1, 2000	
				6. Performing Organization Code AAR-510	
7. Author(s) Susan Monichetti and Michael D. Snyder				8. Performing Organization Report No. DOT/FAA/AR-00/43	
9. Performing Organization Name and Address Federal Aviation Administration William J. Hughes Technical Center Atlantic City International Airport, NJ 08405				10. Work Unit No. (TRAIS)	
				11. Contract or Grant No.	
12. Sponsoring Agency Name and Address Federal Aviation Administration Associate Administrator for Civil Aviation Security, ACS-1 800 Independence Ave., S.W. Washington, DC 20591				13. Type of Report and Period Covered Test & Evaluation Plan	
				14. Sponsoring Agency Code ACS-1	
15. Supplementary Notes This draft test and evaluation plan was prepared by Joshua Rubinstein, Ph.D., Federal Data Corporation.					
16. Abstract This plan outlines a test of Threat Image Projection (TIP) validity. Airport security screeners will evaluate whether fictional threat images (FTIs) are as realistic as actual threat images. The test will compare screener performance on real threat items, FTIs, and Combined Threat Images (CTIs). Performance measures will assess differences in hits, misses, false alarms, and correct rejections for the different types of X-ray images.					
17. Key Words Combined Threat Image, Fictional Threat Image				18. Distribution Statement This report is approved for public release and is on file at the William J. Hughes Technical Center, Aviation Security Research and Development Library, Atlantic City International Airport, New Jersey 08405 This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161	
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of Pages 15	
				22. Price	

TABLE OF CONTENTS

	Page
ACRONYMS	iv
EXECUTIVE SUMMARY	v
1. INTRODUCTION	1
1.1 Background	1
1.2 Scope	1
1.3 Critical Operational Issues and Criteria	1
1.3.1 Issue 1 - Differences in Threat Detection Between Real and TIP Threats	2
1.3.2 Issue 2 - Differences in False Alarm Rates Using Non-Threat Items in Real Bags and TIP Images	2
2. METHOD	2
2.1 Overview of the Test Design	2
2.2 Participants	3
2.3 Test Articles	3
2.4 Procedure	4
2.5 Data Analysis	5
3. TEST LIMITATIONS	6
4. TEST MILESTONES	6
5. REFERENCES	6

LIST OF TABLES

Tables	Page
1 The Distribution of Threat Items in Each of the Three Experimental Sets	3
2 Assignment of Threat Items to the Three Experimental Conditions	4
3 Schedule for Screeners and Scripts	5
4 Test and Evaluation Milestones	6

ACRONYMS

C&E	Compliance and Enforcement
COIC	Critical Operational Issues and Criteria
CTI	Combined Threat Image
FAA	Federal Aviation Administration
FTI	Fictional Threat Image
HFE	Human Factors Engineer
MOP	Measure of Performance
P_d	Probability of Detection
P_{fa}	Probability of False Alarm
TIP	Threat Image Projection
TRX	TIP-Ready X-ray

**PROTECTED UNDER INTERNATIONAL COPYRIGHT
ALL RIGHTS RESERVED
NATIONAL TECHNICAL INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE**

Reproduced from
best available copy.



EXECUTIVE SUMMARY

The Threat Image Projection (TIP) system is an X-ray machine upgrade that projects fictional threats onto or in between the images of passenger baggage. TIP should increase screener vigilance and motivation, provide exposure to a wide range of threat images, and track screener performance. The Federal Aviation Administration's (FAA) Aviation Security Human Factors Program is involved in the deployment, test, and evaluation of TIP at major airports in the United States. In the near future, the FAA will purchase and deploy TIP on more than 400 X-ray machines nationwide. The intent is to use these TIP-ready X-ray (TRX) machines to estimate compliance and enforcement (C&E) of FAA security standards. Voluminous narrative and field data support the validity of TIP for performance estimation, but definitive laboratory validation of the approach is necessary given potential enforcement actions.

This project is an effort to test the validity of TIP by comparing screener performance with TIP images to their performance with real-world threats. The test shall determine if there are any screener performance differences between real threats, projected fictional threat images (FTIs), and projected combined threat images (CTIs). FTIs are fictional threats (e.g., guns, knives, improvised explosive devices, grenades) projected onto passenger baggage as it goes through the X-ray machine. CTIs are fictional bags, which contain threats that are projected in between passenger baggage as it goes through the X-ray machine.

Eighteen TIP-experienced screeners will participate in this study. Each screener will be presented with 510 X-ray images (480 real bags plus 30 CTIs) while operating a TRX machine. Within the 510 X-ray images, there will be 30 fictional threats (15 FTIs and 15 CTIs), 15 real threats, 30 fictional non-threats (15 FTIs and 15 CTIs), and 15 real non-threats. AAR-510 and AAR-540 personnel will create the FTIs and CTIs for each of these 45 threat and 45 non-threat items. Pre-selected bags will also be loaded with these threat and non-threat items. A single script will control the presentation order of the FTIs, CTIs, and real threats. Modified TIP software will be programmed with the script's stimulus presentation order. Normal TIP feedback and any other cues indicating the presence of a TIP image will be disabled.

Screeners will be instructed to operate the X-ray machine as they normally would when screening at a checkpoint. They will either press the TIP "threat" button or allow the bag to pass unchecked. Data analyses will focus on the differences in screener performance during the Real-bag, FTI, and CTI conditions.

1. INTRODUCTION.

The Aviation Security Improvement Act, Public Law 101-604, mandates the Federal Aviation Administration (FAA) to enhance and improve X-ray baggage screener selection, training, and performance. The Aviation Security Human Factors Program (AAR-510) of the Office of Aviation Security Research and Development is the FAA unit tasked with this responsibility.

1.1 BACKGROUND.

The effectiveness of the national civil aviation security system is highly dependent upon the people who are employed as security screeners. Therefore, the FAA is highly interested in the man-machine system operation of new security technologies designed to test, train, improve, and monitor screener performance. The Threat Image Projection (TIP) system is an X-ray machine upgrade that projects fictional threats onto and in between the images of passenger baggage. TIP increases screener vigilance and motivation, provides exposure to a wide range of threat images, and tracks screener performance [1]. AAR-510 is involved in the deployment, test, and evaluation of TIP at major airports in the U.S. [2]. In the near future, the FAA will purchase and deploy TIP on more than 400 X-ray machines nationwide. The intent is to use these TIP-ready X-ray (TRX) machines to estimate compliance with FAA security standards. Voluminous narrative and field data support the validity of TIP for C&E, but definitive laboratory validation of the approach is necessary given potential enforcement actions.

Before these TRX systems are used for performance monitoring and enforcement actions, a laboratory study will be conducted to conclusively demonstrate that fictional threats are as realistic as actual threats to security screeners. Once this has been definitively proven, TIP will be accepted as a valid means for screener and security guard company performance estimation and possibly C&E.

1.2 SCOPE.

This project is an effort to test the validity of TIP by comparing screener performance with TIP images to performance with real-world threats. The test will determine if there are any screener performance differences between real threats, projected fictional threat images (FTIs), and projected combined threat images (CTIs).

1.3 CRITICAL OPERATIONAL ISSUES AND CRITERIA.

The critical operational issues and criteria (COIC) will focus on the differences in Screener performance for different types of threat image presentations. Measures of performance (MOPs) will examine these differences in several ways. Detection of actual threats in real bags will be compared to detection of fictional threats (FTIs and CTIs). Comparisons between performance with FTIs and CTIs will provide valuable information about possible cues provided by the CTI bag itself. An additional comparison of real and TIP images of non-threat items (e.g., a leather shoe) will be performed. This will provide further evidence of differences in performance due to cues provided by the TIP process itself.

1.3.1 Issue 1 - Differences in Threat Detection Between Real and TIP Threats.

Are there differences in a screener's ability to detect threats as FTIs, in CTIs, and in real bags?

Criterion 1-1. There are no differences in threat detection performance between real and TIP images.

MOP 1-1-1. Screeners' mean probability of detection (P_d) for real threats, FTIs, and CTIs.

1.3.2 Issue 2 - Differences in False Alarm Rates Using Non-Threat Items in Real Bags and TIP Images.

Are there differences in a screener's false alarm rates using non-threat items as FTIs, in CTIs, and in real bags?

Criterion 2-1. There are no differences in the percentages of false alarms between real and TIP images.

MOP 2-1-1. The probability of a false alarm (P_{fa}) using non-threat items as FTIs, in CTIs, and in real bags.

2. METHOD.

This plan for TIP validation involves the comparison of the detection of actual threats in real bags, FTIs inserted in real bags, and CTIs. The MOPs associated with the COIC will be collected during the test, and a report describing the results of the test will be created.

2.1 OVERVIEW OF THE TEST DESIGN.

The data collection will take place at the Aviation Security Laboratory at the William J. Hughes Technical Center in Atlantic City, NJ, using experienced airport screeners. Human Factors Engineers (HFEs) will test 18 screeners over a 2-week period. Two screeners will be tested concurrently each day. Screeners will operate an X-ray machine viewing 510 bag images.

This TIP validation study involves a within-subjects design. For each screener, comparisons will be made on TIP performance for threat items presented in real bags, as FTIs, and as CTIs. In addition, non-threat items will be used to provide further data testing for any cues provided by the TIP process.

2.2 PARTICIPANTS.

Eighteen experienced airport security screeners will participate in this study. In order to participate, each screener will be required to have at least 1 month of experience operating TIP on a checkpoint X-ray machine. They must be currently employed as a screener, 18 years of age or older, and not pregnant.

2.3 TEST ARTICLES.

AAR-510 personnel will assemble a collection of 45 threat items. These threats will include 18 improvised explosive devices, 6 guns, 3 grenades, 3 knives, and 15 modular bomb set configurations. These 45 threat items will be divided into three sets. Each threat set will contain equal numbers of each type of threat (e.g., two guns per set). Table 1 shows this distribution of threat items in the three experimental sets. The three sets (A, B, and C) will be assigned to three experimental conditions (real bag, FTI, and CTI). In this way, screener performance for each type of threat can be evaluated when presented in a real bag, as an FTI, or as a CTI.

TABLE 1. THE DISTRIBUTION OF THREAT ITEMS IN EACH OF THE THREE EXPERIMENTAL SETS

	SET A	SET B	SET C
Improvised explosive devices	6	6	6
Guns	2	2	2
Grenades	1	1	1
Knives	1	1	1
Modular bomb sets	5	5	5

AAR-510 personnel will also develop a collection of 45 non-threat items (e.g., a leather shoe). These 45 non-threat items will be assigned to Sets A, B, and C in the same way as the threat items.

AAR-510 personnel will create FTIs and CTIs of each of the 45 threat and 45 non-threat items. A TRX machine will be used to capture these images and store them as FTIs and CTIs in an image library. These images will be used to present the threat and non-threat items to the screener during the experiment.

AAR-510 personnel will gather 480 real bags for this study. For each screener, specific threat and non-threat items will be placed in predetermined bags. The threat-to-bag assignment used for the CTIs will also be used for the real bags during the experiment. This will allow HFEs to examine screener performance differences between TIP and actual threat images without any effects of the bag.

2.4 PROCEDURE.

Rapiscan technicians will modify existing TIP software to allow a controlled presentation order for the TIP images. Normal TIP feedback and any other cues indicating the presence of a TIP image will be disabled. A script will be created that will control the presentation order of the test stimuli. The script will represent the bag order for each of the 480 bags sent through the X-ray machine during the experiment. The script will be programmed into the TIP computer to control the presentation of FTIs and CTIs during the experiment. At predetermined bag-count locations, the script will instruct the computer which FTIs or CTIs to project onto the X-ray monitor. Because CTIs (15 threat and 15 non-threat) include the projection of an entire bag, they will increase the number of test images to 510. The script will also be used by HFEs to control the presentation order of the insertion of the threats into the real bags prior to each experimental run.

It is possible that differences in the identifiability of the threat images could produce unequal performance across conditions. For example, poor screener performance in the FTI condition relative to the Real-bag condition might be because the actual threats used in the FTI condition were harder to identify. In this case, it would be incorrect to infer that there were any differences between the TIP images and real-bag images, per se. To avoid this potential confound, the assignment of specific threat items (Sets A, B, and C) will be counterbalanced across the three conditions using a latin-square design. Table 2 shows the assignment of threat Sets A, B, and C to the FTI, CTI, and Real-bag conditions.

TABLE 2. ASSIGNMENT OF THREAT ITEMS TO THE THREE EXPERIMENTAL CONDITIONS

	FTI	CTI	Real-bag
Screeners 1-6	Item Set A	Item Set B	Item Set C
Screeners 7-12	Item Set B	Item Set C	Item Set A
Screeners 13-18	Item Set C	Item Set A	Item Set B

Because of this counterbalanced design, three different scripts will be developed and programmed into the TIP computer. Six screeners will be assigned to each of the scripts, resulting in 18 screeners tested in this study.

Each screener will be presented with all 510 X-ray images (480 real bags plus 30 CTIs). A single script will control the presentation order of the FTIs, CTIs, and real threats. The modified TIP software will record each time and on which bag the screener presses stop and threat. No feedback will be given during this experiment.

Screeners will be instructed to operate the X-ray machine as they normally would when screening at a checkpoint. There will be two possible responses that screeners can make. They will either press the TIP “threat” button, or they will allow the bag to exit the X-ray process without further examination. Screeners will be able to press the “stop” button as they normally would when screening checkpoint bags. An HFE will record all responses made by the screener during the experiment.

Two X-ray machines will be used, allowing HFEs to test two screeners simultaneously. Each screener will operate the X-ray machine for 30-minute test sessions. They will be provided 15-minute rest intervals between sessions. Using this schedule, screeners should be able to screen all 510 bags within four 30-minute test sessions. Table 3 shows the schedule for screener testing and the assignment of scripts for each test day.

TABLE 3. SCHEDULE FOR SCREENERS AND SCRIPTS

Date	Screeners	Counterbalance
Monday, June 19	1, 2	Script 1
Tuesday, June 20	3, 4	Script 2
Wednesday, June 21	5, 6	Script 3
Thursday, June 22	7, 8	Script 1
Friday, June 23	9, 10	Script 2
Monday, June 26	11, 12	Script 3
Tuesday, June 27	13, 14	Script 1
Wednesday, June 28	15, 16	Script 2
Thursday, June 29	17, 18	Script 3

2.5 DATA ANALYSIS.

HFEs observing and recording screener performance during the test will collect response data. For each bag, they will record whether the screener pressed the “threat” or “stop” button or let the bag pass. The proportion of hits, misses, false alarms, and correct rejections will be calculated. Analyses of variance will be performed on the data to determine if there are any differences in screener performance across experimental conditions.

3. TEST LIMITATIONS.

One possible limitation of this validation test concerns the generalizability of the results to other TIP systems. Only two machines from one company (Rapiscan) will be used in this test. TIP software produced by other companies may be sufficiently different that it could produce TIP images that cue screeners as to the veracity of the X-ray image. One way to attenuate this potential limitation is to repeat this evaluation with X-ray machines produced by other companies.

A second limitation associated with this study concerns the fact that a lack of statistical significance between conditions is an important result. Looking for experimental conditions that produce no statistical differences in screener performance is referred to as “proving the null hypothesis.” This is not a methodologically and statistically strong position to take. It is always possible that insensitivities in the experimental design or implementation contribute to the reasons for not finding statistical differences between conditions. One way to attenuate this potential limitation is to replicate this result under different conditions and using different stimuli. In addition, statistical procedures (e.g., equivalency testing) can be used that may be more appropriate for this type of study.

4. TEST MILESTONES.

Table 4 shows the milestones for planning and reporting the operational test & evaluation. FDC will be responsible for preparing all reports.

TABLE 4. TEST AND EVALUATION MILESTONES

Milestone	Date
Test and Evaluation Plan	March 20, 2000
Laboratory Test Preparation	June 5 to June 16, 2000
Laboratory Test and Evaluation	June 19 to June 30, 2000
Preliminary Report	Within 15 days of completion of testing
Final Report	July 28, 2000

5. REFERENCES.

1. Neiderman, E. C., Fobes, J. L., Klock, B. A., & Barrientos, J. M., “Functional Requirements for Threat Image Projection Systems on X-Ray Machines” (DOT/FAA/AR-97/67), DOT/FAA William J. Hughes Technical Center, Atlantic City International Airport, NJ, 1997.

2. Neiderman, E. C., "Test and Evaluation Plan for Airport Demonstration of Threat Image Projection for Checkpoint Operations" (DOT/FAA/AR-96/92), DOT/FAA William J. Hughes Technical Center, Atlantic City International Airport, NJ, 1996.

